# Quantitative Assessment of Bias Sensitivity of Performance Measures for Dichotomous Forecasts

Keith F. Brill NCEP/HPC







#### Reference



This presentation is based mostly on a paper accepted for publication in *Weather and Forecasting*, now available on line at the American Meteorological Society Publications page under the link "Early Online Releases of Papers in Press (2008)." Select the link for *Weather and Forecasting*, and look for the posting date of July 23, 2008.



## Overview



- Review definitions and 2X2 contingency table
- State motivations and goals
- Derive a Critical Performance Ratio (CPR) that quantifies bias sensitivity
- Apply the CPR to reveal bias dependencies of several performance metrics
- Summarize and discuss future work

## **Definitions**

- Dichotomous forecasts: "yes" or "no" forecasts for occurrence of some event, e.g., precipitation accumulation exceeding a threshold
- **Bias** (*B*, **frequency bias**): the ratio of number or frequency of "yes" forecasts to "yes" observations
- Probability of Detection (P, POD): the ratio of number or frequency of correct "yes" forecasts to "yes" observations
- **Event frequency**  $(\alpha)$ : the fraction of the entire verification domain (temporal and spatial) comprised of "yes" observations

# Contingency Table

EVENTS	Observed	Not Observed	Total
Forecast	H	F-H	F
Not Forecast	O–H	N-F-O+H	N–F
Total	0	N-O	N

#### Where:

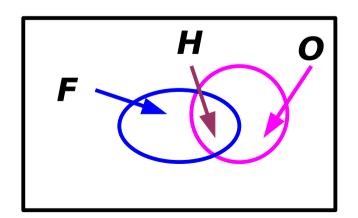
H = number or area of correct "yes" forecasts, hits

F = number or area of "yes" forecasts

O = number or area of "yes" observations

N = total number or area constituting the verification domain

$$B = \frac{F}{O} \qquad P = \frac{H}{O} \qquad \alpha = \frac{O}{N} > 0$$



## **Motivations and Goals**

- MOTIVATIONS: Performance measures computed from F, H, O values are known from experience to be sensitive to bias (e.g., Baldwin and Kain, 2006), having implications for
  - Assessing "hedged" forecasts
  - Assessing bias correct forecasts
  - Assessing forecasts evaluated using Spatial Techniques

#### GOALS:

- Derive a general mathematical expression quantifying bias sensitivity
- Apply the general quantitative expression to specific performance measures

# Analytical Approach

- Rewrite contingency table in terms of P, B, and  $\alpha$
- Express a performance measure in terms of P, B, and  $\alpha$  as independent variables
- Assume verification (retrospective) point of view so that  $\alpha$  may be considered constant
- Use total derivative of a performance measure to determine how P must change with B for the performance measure to indicate improvement

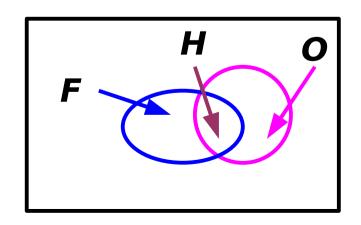
# Rewrite Contingency Table

EVENTS	Observed	Not Observed	Total	
Forecast	H	F-H	F	
Not Forecast	O–H	N-F-O+H	N-F	
Total	0	N-O	N	

Multiply each cell by O/O=1. Normalize by dividing by N. Replace H/O with P, F/O with B, and O/N with  $\alpha$ .

EVENTS	Observed	Not Observed	Total
Forecast	$\alpha P$	$\alpha(B-P)$	$\alpha B$
Not Forecast	$\alpha(1-P)$	$1-\alpha (B+1-P)$	$1-\alpha B$
Total	α	$1-\alpha$	1

$$B = \frac{F}{O} \qquad P = \frac{H}{O} \qquad \alpha = \frac{O}{N} > 0$$



# Mathematical Analysis

Write any performance measure as S=S(B,P), with  $\alpha$  constant.

Express total differential of *S*: 
$$dS = (\frac{\partial S}{\partial B})_P dB + (\frac{\partial S}{\partial P})_B dP$$
. (1)

Consider **positively oriented** performance measures that indicate improvement by increasing in value. Thus, for small changes in B and P, S indicates improvement if

$$\left(\frac{\partial S}{\partial B}\right) \Delta B + \left(\frac{\partial S}{\partial P}\right) \Delta P > 0.$$
 (2)

It follows that

$$\left(\frac{\partial S}{\partial P}\right) \Delta P > -\left(\frac{\partial S}{\partial B}\right) \Delta B. \tag{3}$$

Since  $H=\alpha P$  and  $F=\alpha B$ , if  $\alpha$  is constant, then  $\Delta P=\Delta H/\alpha$  and  $\Delta B=\Delta F/\alpha$ .

# Mathematical Analysis Continued

$$\left(\frac{\partial S}{\partial P}\right) \frac{\Delta H}{\alpha} > -\left(\frac{\partial S}{\partial B}\right) \frac{\Delta F}{\alpha} . \tag{4}$$

Consider an increase in bias,  $\Delta F > 0$ , assuming  $(\partial S/\partial P) > 0$ , then the general condition for S increase is

$$\frac{\Delta H}{\Delta F} > -\frac{\left(\frac{\partial S}{\partial B}\right)}{\left(\frac{\partial S}{\partial P}\right)} = \rho. \tag{5}$$

For a decrease in bias,  $\Delta F < 0$ , still assuming  $(\partial S/\partial P) > 0$ , the general condition for S increase is

$$\frac{\Delta H}{\Delta F} < \rho. \tag{6}$$

# Mathematical Analysis Continued

$$\rho = -\frac{(\frac{\partial S}{\partial B})}{(\frac{\partial S}{\partial P})}$$
 defines the critical performance ratio (CPR). (6)

In summary, if bias is increased, a performance measure indicates improvement if

hit fraction for added forecasts = 
$$\frac{\Delta H}{\Delta F} > \rho$$
. (7)

For a decrease in bias, a performance measure indicates improvement if

hit fraction of removed forecasts = 
$$\frac{\Delta H}{\Delta F} < \rho$$
. (8)

The same conditions for *S* to improve obtain for negatively oriented performance measures.

Ineq. (7) or (8) expresses the CPR criterion for a performance measure to improve for a change in bias.





For a decrease in bias, the CPR sets a bar to get under.

For an increase in bias, the CPR sets a bar to get over.

# Mathematical Analysis Concluded

The following three constraints apply:

$$1. \quad 0 \leq \frac{\Delta H}{\Delta F} \leq 1$$

Additional (removed) hits cannot exceed the change in forecast number or area for an increase (decrease) in bias.

2. 
$$\frac{\Delta H}{\Delta F} = 0$$
 If  $P=1$  and  $\Delta B > 0$ 

3. 
$$\frac{\Delta H}{\Delta F} = 0$$
 If  $P=0$  and  $\Delta B < 0$ 

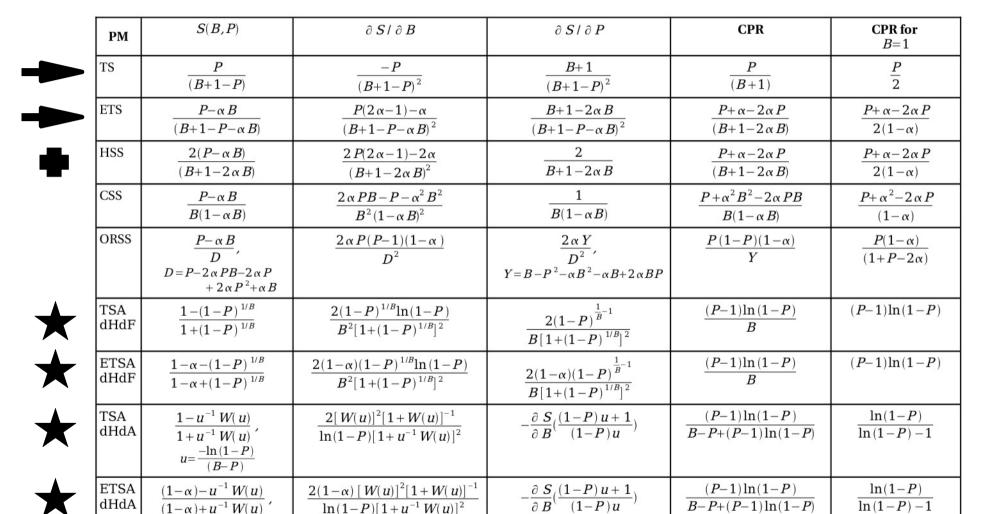
If the CPR condition requires a violation of one of these constraints, the performance measure cannot indicate improvement.

# **Analysis Method Summary**

- 1. Express *S* in terms of *B*, *P*, and  $\alpha > 0$ .
- 2. Derive and simplify required partial derivatives.
- 3. Evaluate  $(\partial S/\partial P)$  to assure correct algebraic sign.
- 4. Compute the CPR,  $\rho$ .
- 5. Select the appropriate inequality based on the sign of the bias change.
- 6. If the CPR criterion violates any one of the three constraints, the performance measure cannot indicate improvement.

CPRs may be derived and examined graphically as functions of B, P, and/or  $\alpha$ .

## Table of Derivative & CPR Formulas for Selected Performance Measures



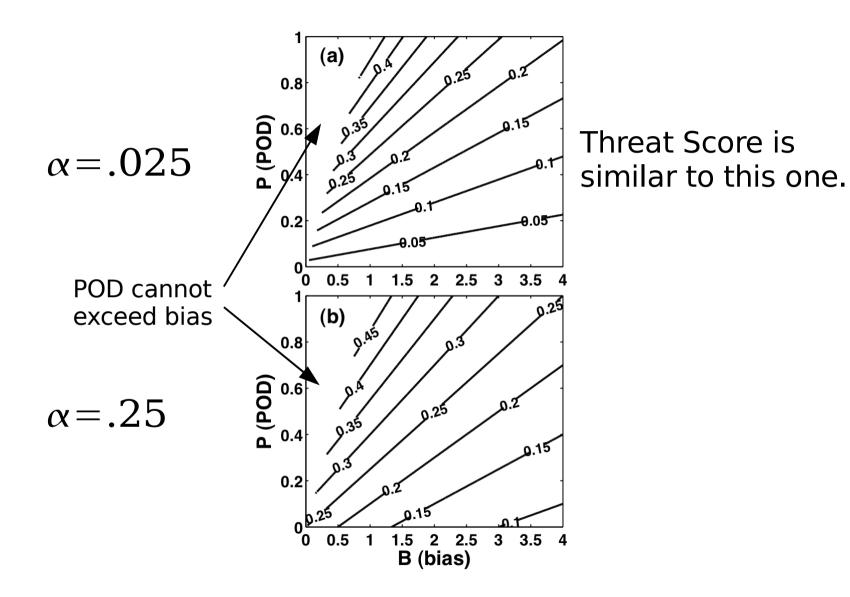


Mesinger (2008) bias adjusted TS & ETS

 $u = \frac{-\ln(1-P)}{(B-P)}$ 

**Lambert W function:**  $z = W(z) e^{W(z)}$ .  $1 = W(z) \frac{dW(z)}{dz} e^{W(z)} + \frac{dW(z)}{dz} e^{W(z)}$ .  $\frac{dW(z)}{dz} = \frac{W(z)}{z \lceil 1 + W(z) \rceil}$ .

#### ETS CPR Contours on the POD-Bias Plane



## Table of Derivative & CPR Formulas for Selected Performance Measures

PM	S(B,P)	∂ S / ∂ B	∂ S / ∂ P	CPR	CPR for B=1
TS	$\frac{P}{(B+1-P)}$	$\frac{-P}{(B+1-P)^2}$	$\frac{B+1}{(B+1-P)^2}$	$\frac{P}{(B+1)}$	$\frac{P}{2}$
ETS	$\frac{P\!-\!\alphaB}{(B\!+\!1\!-\!P\!-\!\alphaB)}$	$\frac{P(2\alpha-1)-\alpha}{(B+1-P-\alpha B)^2}$	$\frac{B+1-2\alpha B}{(B+1-P-\alpha B)^2}$	$\frac{P+\alpha-2\alpha P}{(B+1-2\alpha B)}$	$\frac{P+\alpha-2\alpha P}{2(1-\alpha)}$
HSS	$\frac{2(P\!-\!\alphaB)}{(B\!+\!1\!-\!2\alphaB)}$	$\frac{2P(2\alpha-1)-2\alpha}{(B+1-2\alpha B)^2}$	$\frac{2}{B+1-2\alpha B}$	$\frac{P+\alpha-2\alpha P}{(B+1-2\alpha B)}$	$\frac{P+\alpha-2\alpha P}{2(1-\alpha)}$
CSS	$\frac{P-\alphaB}{B(1-\alphaB)}$	$\frac{2\alpha PB - P - \alpha^2 B^2}{B^2 (1 - \alpha B)^2}$	$\frac{1}{B(1-\alpha B)}$	$\frac{P+\alpha^2B^2-2\alpha PB}{B(1-\alpha B)}$	$\frac{P+\alpha^2-2\alpha P}{(1-\alpha)}$
ORSS	$\frac{P-\alpha B}{D},$ $D=P-2\alpha PB-2\alpha P$ $+2\alpha P^{2}+\alpha B$	$\frac{2\alpha P(P-1)(1-\alpha)}{D^2}$	$\frac{2\alpha Y}{D^{2}},$ $Y = B - P^{2} - \alpha B^{2} - \alpha B + 2\alpha BP$	$\frac{P(1-P)(1-lpha)}{Y}$	$\frac{P(1-\alpha)}{(1+P-2\alpha)}$
TSA dHdF	$\frac{1 - (1 - P)^{1/B}}{1 + (1 - P)^{1/B}}$	$\frac{2(1-P)^{1/B}\ln(1-P)}{B^2[1+(1-P)^{1/B}]^2}$	$\frac{2(1\!-\!P)^{\frac{1}{B}\!-\!1}}{B[1\!+\!(1\!-\!P)^{1/B}]^{2}}$	$\frac{(P\!-\!1)\!\ln{(1\!-\!P)}}{B}$	$(P-1)\ln(1-P)$
ETSA dHdF	$\frac{1 - \alpha - (1 - P)^{1/B}}{1 - \alpha + (1 - P)^{1/B}}$	$\frac{2(1-\alpha)(1-P)^{1/B}\ln(1-P)}{B^2[1+(1-P)^{1/B}]^2}$	$rac{2(1\!-\!lpha)(1\!-\!P)^{rac{1}{B}\!-1}}{B[1\!+\!(1\!-\!P)^{1/B}]^2}$	$\frac{(P-1)\ln{(1-P)}}{B}$	$(P-1)\ln(1-P)$
TSA dHdA	$\frac{1 - u^{-1} W(u)}{1 + u^{-1} W(u)},$ $u = \frac{-\ln(1 - P)}{(B - P)}$	$\frac{2[W(u)]^{2}[1+W(u)]^{-1}}{\ln(1-P)[1+u^{-1}W(u)]^{2}}$	$-\frac{\partial}{\partial} \frac{S}{B} (\frac{(1-P)u+1}{(1-P)u})$	$\frac{(P-1)\ln(1-P)}{B-P+(P-1)\ln(1-P)}$	$\frac{\ln(1-P)}{\ln(1-P)-1}$
ETSA dHdA	$\frac{(1-\alpha)-u^{-1} W(u)}{(1-\alpha)+u^{-1} W(u)},$ $u = \frac{-\ln(1-P)}{(B-P)}$	$\frac{2(1-\alpha) [W(u)]^{2} [1+W(u)]^{-1}}{\ln(1-P)[1+u^{-1} W(u)]^{2}}$	$-\frac{\partial}{\partial} \frac{S}{B} \left(\frac{(1-P)u+1}{(1-P)u}\right)$	$\frac{(P-1)\ln(1-P)}{B-P+(P-1)\ln(1-P)}$	$\frac{\ln(1-P)}{\ln(1-P)-1}$

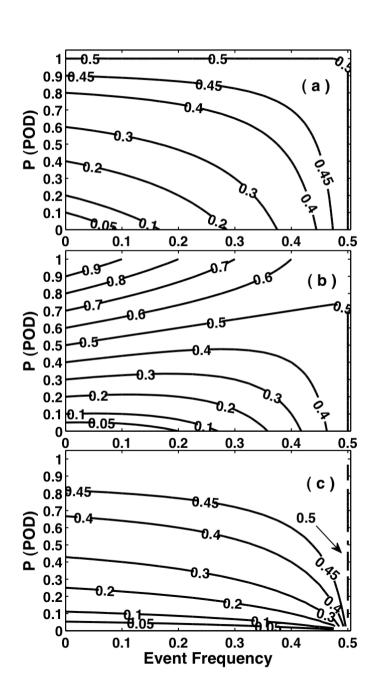
**Lambert W function:**  $z = W(z) e^{W(z)}$ .  $1 = W(z) \frac{dW(z)}{dz} e^{W(z)} + \frac{dW(z)}{dz} e^{W(z)}$ .  $\frac{dW(z)}{dz} = \frac{W(z)}{z[1 + W(z)]}$ .

#### CPR Contours on the POD- $\alpha$ Plane for B=1

Equitable Threat Score

Clayton Skill Score

Odds Ratio Skill Score



# Mesinger (2008) Bias Adjusted TS & ETS

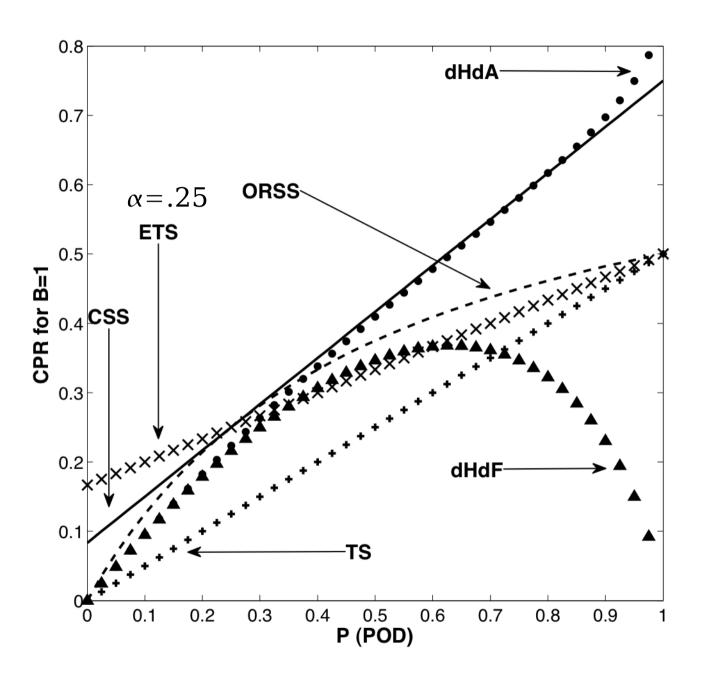
- Uses F, H, O values to interpolate/extrapolate H to the condition B=1, with  $H_a$  hits.
- Presents two methods for computing H<sub>a</sub>:
  - dHdF assumes hit area change with respect to forecast area is proportional to (O-H).
  - dHdA assumes hit area change with respect to false alarmed area is proportional to (O-H).
- Computes TS or ETS using  $H_a$  and F=O (B=1) to account for errors in placement.

## Table of Derivative & CPR Formulas for Selected Performance Measures

PM	S(B,P)	∂ S / ∂ B	∂ S / ∂ P	CPR	CPR for B=1
TS	$\frac{P}{(B+1-P)}$	$\frac{-P}{(B+1-P)^2}$	$\frac{B+1}{(B+1-P)^2}$	$\frac{P}{(B+1)}$	$\frac{P}{2}$
ETS	$\frac{P-\alpha B}{(B+1-P-\alpha B)}$	$\frac{P(2\alpha-1)-\alpha}{(B+1-P-\alpha B)^2}$	$\frac{B+1-2\alpha B}{(B+1-P-\alpha B)^2}$	$\frac{P+\alpha-2\alpha P}{(B+1-2\alpha B)}$	$\frac{P+\alpha-2\alpha P}{2(1-\alpha)}$
HSS	$\frac{2(P-\alpha B)}{(B+1-2\alpha B)}$	$\frac{2P(2\alpha-1)-2\alpha}{(B+1-2\alpha B)^2}$	$\frac{2}{B+1-2\alpha B}$	$\frac{P+\alpha-2\alpha P}{(B+1-2\alpha B)}$	$\frac{P+\alpha-2\alpha P}{2(1-\alpha)}$
CSS	$\frac{P-\alpha B}{B(1-\alpha B)}$	$\frac{2\alpha PB - P - \alpha^2 B^2}{B^2 (1 - \alpha B)^2}$	$\frac{1}{B(1-\alpha B)}$	$\frac{P+\alpha^2B^2-2\alpha PB}{B(1-\alpha B)}$	$\frac{P+\alpha^2-2\alpha P}{(1-\alpha)}$
ORSS	$\frac{P-\alpha B}{D},$ $D=P-2\alpha PB-2\alpha P$ $+2\alpha P^{2}+\alpha B$	$\frac{2\alpha P(P-1)(1-\alpha)}{D^2}$	$\frac{2\alpha Y}{D^{2}},$ $Y = B - P^{2} - \alpha B^{2} - \alpha B + 2\alpha BP$	$\frac{P(1-P)(1-lpha)}{Y}$	$\frac{P(1-\alpha)}{(1+P-2\alpha)}$
TSA dHdF	$\frac{1 - (1 - P)^{1/B}}{1 + (1 - P)^{1/B}}$	$\frac{2(1-P)^{1/B}\ln(1-P)}{B^2[1+(1-P)^{1/B}]^2}$	$\frac{2(1\!-\!P)^{\frac{1}{B}\!-\!1}}{B[1\!+\!(1\!-\!P)^{^{1/B}}]^{^2}}$	$\frac{(P-1)\ln{(1-P)}}{B}$	$(P-1)\ln(1-P)$
ETSA dHdF	$\frac{1-\alpha-(1-P)^{1/B}}{1-\alpha+(1-P)^{1/B}}$	$\frac{2(1-\alpha)(1-P)^{1/B}\ln{(1-P)}}{B^2[1+(1-P)^{1/B}]^2}$	$rac{2(1\!-\!lpha)(1\!-\!P)^{rac{1}{B}\!-\!1}}{B[1\!+\!(1\!-\!P)^{1/B}]^2}$	$\frac{(P-1)\ln{(1-P)}}{B}$	$(P-1)\ln(1-P)$
TSA dHdA	$\frac{1 - u^{-1} W(u)}{1 + u^{-1} W(u)},$ $u = \frac{-\ln(1 - P)}{(B - P)}$	$\frac{2[W(u)]^{2}[1+W(u)]^{-1}}{\ln(1-P)[1+u^{-1}W(u)]^{2}}$	$-\frac{\partial}{\partial} \frac{S}{B} (\frac{(1-P)u+1}{(1-P)u})$	$\frac{(P-1)\ln(1-P)}{B-P+(P-1)\ln(1-P)}$	$\frac{\ln(1-P)}{\ln(1-P)-1}$
ETSA dHdA	$\frac{(1-\alpha)-u^{-1} W(u)}{(1-\alpha)+u^{-1} W(u)},$ $u = \frac{-\ln(1-P)}{(B-P)}$	$\frac{2(1-\alpha) [W(u)]^{2} [1+W(u)]^{-1}}{\ln(1-P)[1+u^{-1} W(u)]^{2}}$	$-\frac{\partial}{\partial} \frac{S}{B} (\frac{(1-P)u+1}{(1-P)u})$	$\frac{(P-1)\ln(1-P)}{B-P+(P-1)\ln(1-P)}$	$\frac{\ln(1-P)}{\ln(1-P)-1}$

Lambert W function:  $z = W(z) e^{W(z)}$ .  $1 = W(z) \frac{dW(z)}{dz} e^{W(z)} + \frac{dW(z)}{dz} e^{W(z)}$ .  $\frac{dW(z)}{dz} = \frac{W(z)}{z[1 + W(z)]}$ .

#### CPR vs POD for B=1



# Applications

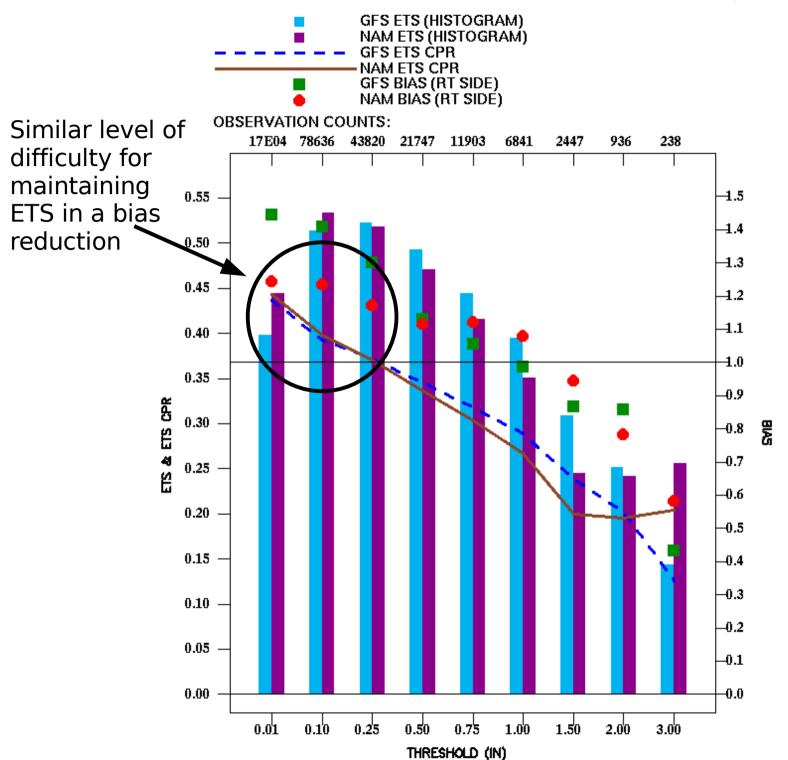
- ETS and DHDA ETS CPRs vs threshold for cold season NAM and GFS QPF.
- ETS and DHDA ETS CPRs vs threshold for warm season NAM and GFS QPF.

Note to fvs users: Version 2008.07 of fvs computes and

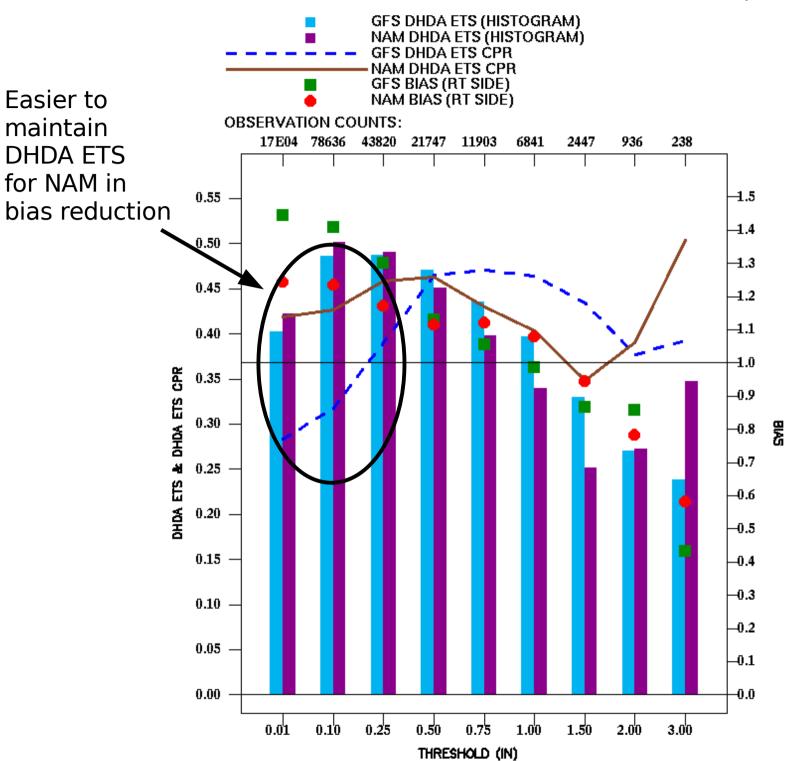
displays CPR functions for all performance measures for FHO stats. Enter "fvs v" for more information on computational codes

1036--1067.

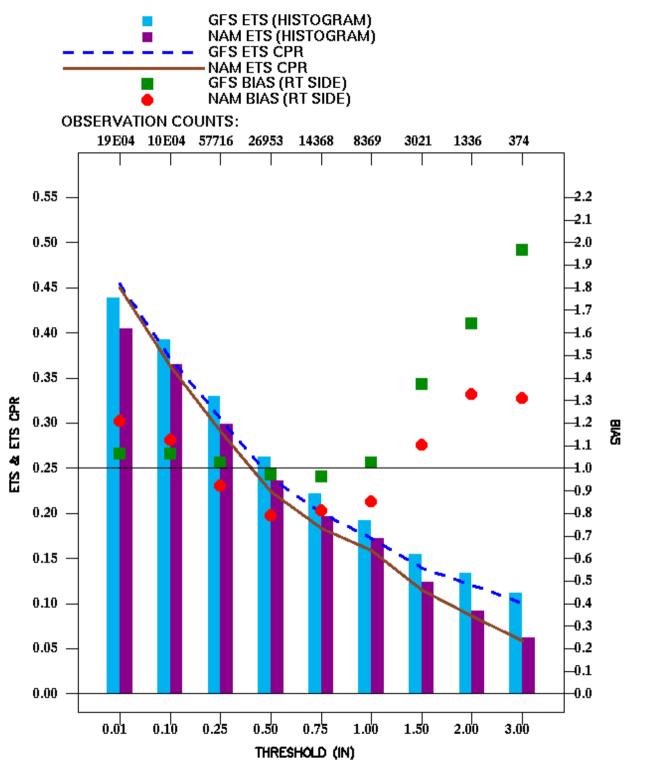
#### DJF 07-08 212/RFC PERFORMANCE FOR 24-H FORECAST OF 24-H QPF



#### DJF 07-08 212/RFC PERFORMANCE FOR 24-H FORECAST OF 24-H QPF



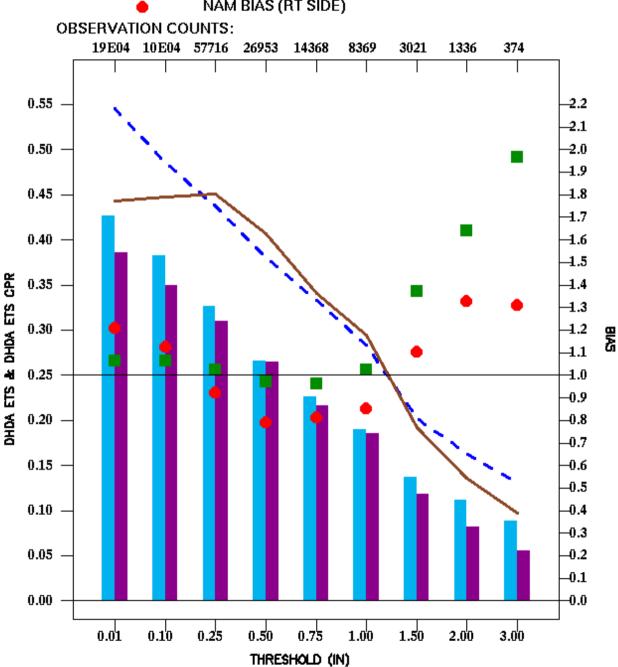
#### JJA 2008 212/RFC PERFORMANCE FOR 24-H FORECAST OF 24-H QPF



#### JJA 2008 212/RFC PERFORMANCE FOR 24-H FORECAST OF 24-H QPF



DHDA ETS is more resistant to showing improvement if hedging inflates bias.



# Summary

- The critical performance ratio CPR quantifies bias sensitivity.
- The CPR and, therefore, bias sensitivity depend on one or more of POD, bias, and event frequency.
- All performance measures are sensitive to bias.
- CPR values for Mesinger's bias adjusted ETS may give a better indication of how easily ETS can be maintained or improved in a bias correction, especially if bias is large.
- DHDA ETS is more resistant to showing improvement if hedging inflates bias.

## **Future Work**

- Publish a note on the CPR for bias adjusted TS and ETS.
- Investigate two bench marks for the CPR
  - Two conditional probabilities have CPR expressions identical to themselves:
    - Detection Failure Ratio = the chance of randomly making hits for an increase in forecast area
    - Frequency of Hits (post agreement) = the chance of randomly loosing hits for a decrease in forecast area

#### JJA 2008 212/RFC GFS 84-H FORECAST OF 24-H QPF CPR FOR THREAT SCORE **CPR FOR EQ. THREAT SCORE** CPR FOR DHDA EQ. TS DH/DF FOR RANDOM INCREASE IN BIAS DH/DF FOR RANDOM DECREASE IN BIAS BIAS (RIGHT SIDE) **OBSERVATION COUNTS:** 18E04 94653 53047 24929 7826 2892 1290 372 13303 0.70 -1.4-1.3 0.650.60-1.2 0.55 -1.1E 0.50-1.0H 0.45 -0.9 CPR AND RANDOM DH/DF 0.40 -0.8 E -0.7 0.35 0.30 -0.6 0.25 -0.50.20 -0.40.15 --0.3 -0.2 0.10 -0.1 0.050.00-0.0 29 0.251.00 0.10 0.50 0.75 1.50 2.00 0.013.00 THRESHOLD (IN)